Authors' response

Past, Present, and Future Variability of Atlantic Meridional Overturning Circulation in CMIP6 Ensembles

We would like to thank the reviewers for the helpful comments and suggestions. They helped us to improve our manuscript considerably, especially the comments on the interpretation of interactions which led to further work and helped us to better understand the physical significance of these interactions.This enhanced understanding allows us to introduce two additional sections: one detailing the interpretation of interactions (Section 2.2.2), and another presenting a new method for combining the ANOVA components (Section 3.4). In addition, this new interpretation has shown us that Fig. 4 in the first version of the manuscript should be based on main effects and interactions instead of just the main effects. The updated figure is presented at the end of this response and does not change the interpretation of the results.
Finally, we have responded to each reviewer's comment below and provide a revised version of the manuscript.

Arthur Coquereau, Florian Sévellec, Thierry Huck, Joël J.-M. Hirschi, and Quentin Jamet

Comments from the editor

This manuscript offers a detailed analysis of the Atlantic Meridional Overturning Circulation (AMOC) using CMIP6 multi-model ensembles, applying an ANOVA-based decomposition to separate internal, model, scenario, and temporal components of AMOC variability from 1850 to 2100. It identifies three distinct variability regimes and finds that internal variability decreases alongside AMOC intensity, a result seemingly proportional to emission scenario intensity. Model-related uncertainty is identified as the dominant contributor throughout the historical and projected periods. The topic is both timely and impactful, as AMOC is a major driver of climate variability and long-term projections. However, while the work is methodologically structured, the paper lacks critical depth in terms of physical interpretation, especially in light of emerging insights from high-resolution models, which reveal behavior not captured in the CMIP6 standard resolution framework. The interplay between convection, eddy activity, and AMOC is scale dependent, posing a challenge for the large-scale circulation and mesoscale features in a warming ocean. I suggest a deeper analysis or at least a discussion of the different branches of the AMOC and convection sites (c.f. Gou et al., 2024, PRL). Furthermore, please follow the suggestions from the referees.

We thank the editor for his/her feedback and agree concerning the lack of physical interpretation, as also noted by reviewer #2 in comments 1, 5 and 7. We added a full subsection on interpretation in the methodology section as described in detail in the response to comment #5 of the second reviewer.

Moreover, model resolution remains an important constraint in the analysis of AMOC time series from climate models. As suggested, this caveat have been added in the discussion section:

"*A limitation of this study, tied to modeling capacity, lies in the relatively small number of large ensemble climate models and the limited size of most ensembles. This constraint is evident, for example, in the analysis of ensemble variance within individual models (Fig. 6), where variability exhibits noisy behavior and notable differences across small ensemble models, despite a generally consistent pattern of decline. Additionally, while some high-resolution climate simulations for a single member are becoming available, large-scale ensemble simulations remain restricted to relatively coarse spatial resolutions – on the order of one degree in the ocean. In this context, different research groups have highlighted the importance of fine-scale processes, such as mesoscale eddies, overflows and convection, in influencing the AMOC, raising concerns about the reliability of AMOC trends in coarse-resolution (one-degree) models (e.g., Hirschi et al., 2020; Hewitt et al., 2022; Gou et al., 2024). However, regarding AMOC strength at 26°N, a recent study found strong agreement between high- and low-resolution simulations (with 0.1° resolution in the ocean, Gou et al., 2024), which is the central focus of this work.*" (l. 530-539)

Comments from reviewer #1

This study investigates the variability of AMOC across scenarios, models, ensembles and time in CMIP6 models. The authors applied a novel method, and this part needs other reviewers' expertise. It is a solid study, but it is difficult to see the key messages, and the implications and discussions of the results should be improved. At least the key points highlighted in the end should be better discussion in the middle sections. Some details are needed – for example, please elaborate on the initial conditions of the model ensemble (L51-53, Table 1), and the reason for incorporating the time dimension is unclear (L71-72).

We thank the reviewer for the feedback. We have addressed each point and provide a revised version of the manuscript:

1. Key Messages

We understand the concern of the reviewer about the visibility of our key messages. We have revised the manuscript to better highlight our main findings by introducing the key points from the "Summary and Discussion" directly throughout the “Results” section.

2. Initial Conditions

We agree that further explanation of the initial conditions of the ensemble models is relevant. We added the following paragraph in the material section:

“***Here, we focus on initial-condition ensemble simulations to sample the phase-space and investigate the spread of the different possible trajectory as a proxy of internal variability (Fig. 1). Initial conditions are derived following a predefined strategy for the CMIP6 framework, starting with different years of the multi-secular preindustrial control run (known as piControl), which is run under fixed external forcing conditions from the year 1850 (Eyring et al., 2016).***“ (l. 87-91)

3. Time Dimension

We agree with the reviewer that the introduction lacks an important motivation for incorporating the time dimension: the simple fact that adding the time dimension allows us to study how the temporal variability of the system changes over time. It allows us to study interannual-to-decadal variability through successive 30-year climate periods and estimate the evolution of this interannual-to-decadal variability (including trends). As an example, when the AMOC declines in CMIP6, this variation is a factor of variability for AMOC. However, when time variability is not taken into account, this contribution is not visible. Historically ANOVA has been applied without incorporating the time dimension.This can probably be explained by the fact that ANOVA is often used to characterize uncertainty, and that a trend common to all scenarios, models, and members is not a factor of uncertainty, whereas it is a factor of variability. Hence incorporating the time dimension allows us to generalize ANOVA methodology for variability study. This was mentioned in the "Method" section (previously l. 111-112), but we explicitly added this information in the introduction:

"*In this study, we propose incorporating the time dimension to examine how the interannual-to-decadal variability of the AMOC – including long-term trends – has changed over time, by analyzing successive 30-year climate periods. Previous climate studies using ANOVA did not include time because the method was mainly used to measure uncertainty. A trend that is common across all scenarios, models, and ensemble members does not contribute to uncertainty per se, but rather to variability. By explicitly including time, we generalize the ANOVA framework to study not just uncertainty, but how variability itself changes over time. Accounting for time variability is particularly critical in the historical period where there are no forcing scenarios; in such cases, neglecting time dimension prevents the detection of changes in the climate state driven by external factors.*" (l. 70-77)

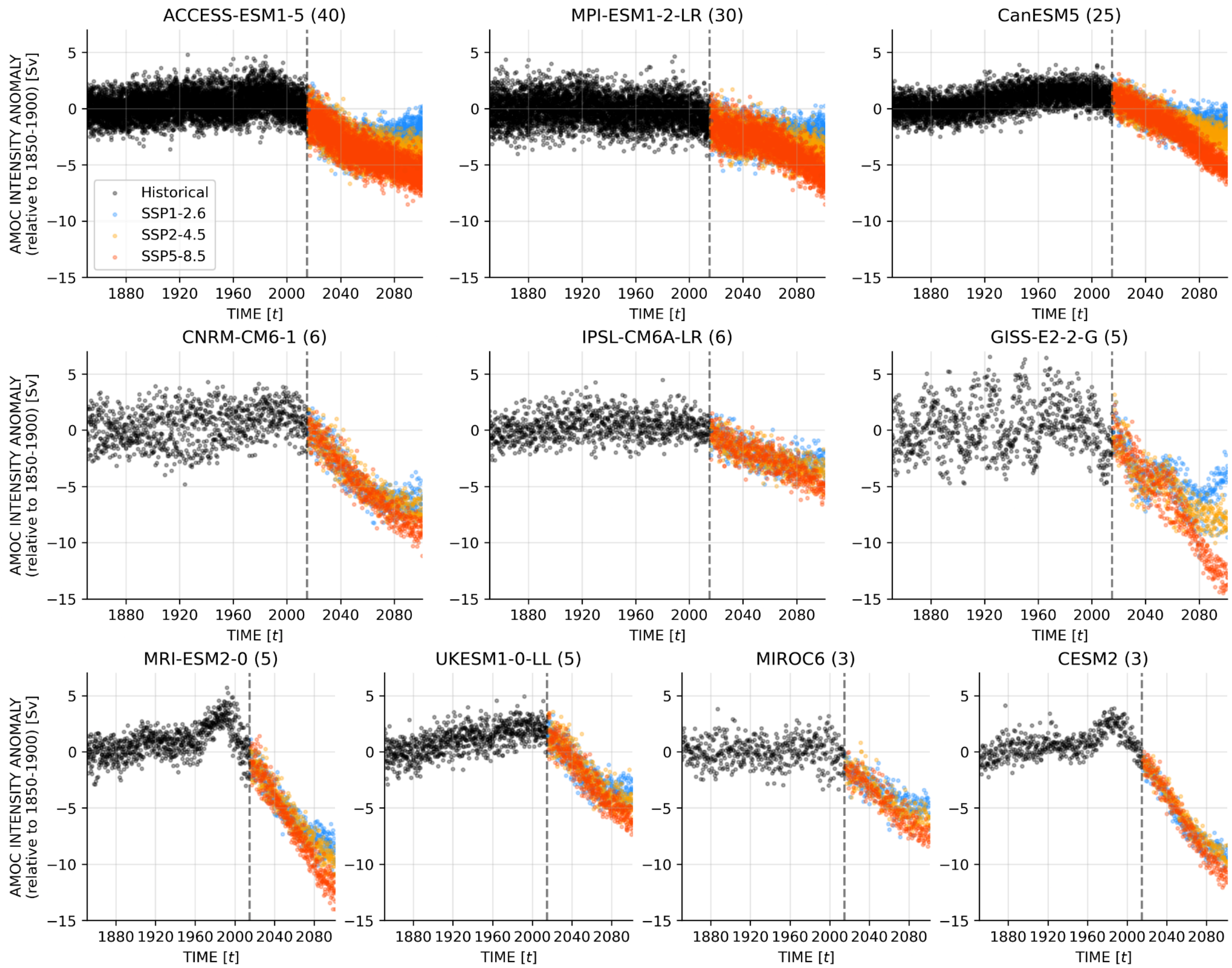
References:

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Comments from reviewer #2

1. I miss an interpretation of what I see in Fig. 1. The caption explains what each curve represents, but how to interpret the differences is unclear to me and not (well) discussed in the text.

We agree with the reviewer about the lack of interpretation. This figure is mainly qualitative, illustrating what is involved in averaging a dimension. However, upon revision, we felt that the figure could be misleading, as it introduced an alternative perspective on variability that might confuse the reader. As a result, we have decided to remove the figure, as it no longer seems essential. Instead, we have included a simpler figure showing the time series from each ensemble member/realization (see below). This provides a more straightforward overview of the data and offers helpful context for the ANOVA decomposition.



Caption: **Evolution of AMOC intensity anomaly.** Each subplot presents historical time series (black) followed by the three scenarios SSP1-2.6 (blue), SSP2-4.5 (orange), or SSP5-8.5 (red) for a given model. For each model, all realizations/members of the ensemble are shown, and the relative intensity uses as a reference the ensemble mode average between 1850 and 1900.

2. In the same Fig. different y-axes are used, which is misleading. I would suggest to you use the same y-scale everywhere (the one from panel f), or at least the panel f scale in b,d,f and the panel e scale in a,c,e.

As explained in the reply to comment #1, we have removed this figure from the revised manuscript.

3. I think it is misleading to talk about the “large ensembles” and the “small ensembles” in many places, suggesting that differences in AMOC variability are due to ensemble size, while it is most probably to different models being analysed and more and less models available in the small and large ensemble. I suggest to use “The models with a large ensemble” vs “the models with a smaller ensemble”, or just ensemble A and ensemble B. This especially holds to the amount of decline. The much larger decline in the small ensemble is simply due to including models with larger AMOC sensitivity while the large ensemble by chance consists of 3 models that do not weaken much. This fact has nothing to do with ensemble size. On the other hand, when discussing the interactions I accept that ensemble size can be driving factor for differences in certain periods.

We agree that some of the differences between small and large ensembles are not directly due to the size of the ensembles, but rather to the characteristics of the models, as suggested by the reviewer with the sensitivity of AMOC to forcing. This aspect of sensitivity is probably reflected in the factors and interactions associated with the time and scenario dimensions. The factors and interactions associated with scenarios behave relatively similarly between small and large ensembles. Qualitatively, this is also the case for the main effect of time, but not quantitatively. The greater amplitude of the time main effect (T) bump for small ensembles and its greater spread over time could be due to the sensitivity of AMOC to forcing. However, this may also be linked to the small size of the ensembles, which does not allow internal variability to be separated as precisely as for larger ensembles. The reviewer’s suggestion also provides a convincing explanation for the stronger model-time interaction (MT) in small ensembles, as well as for the model main effect (M). We added these arguments in the revised manuscript (see below). We note, however, that the separation between small and large ensembles remains justified in view of the factors associated with internal variability. Finally, we agree with the reviewer concerning the suggestion of defining clearly the two distinct categories of models. We have defined the terms “large ensemble models” and “small ensemble models” in the “Material” section of the revised manuscript and used them throughout the text.

“***This is not due to any inherent characteristic of small ensembles, but rather associated with the particular combination of models in the "small ensemble" and "large ensemble" category, which happen to exhibit different AMOC sensitivities to forcing***” (l. 448-450)

4. Please say what R T, S mean in the caption of Fig. 3. Figures should be stand-alone understandable without having to go back to the text to see what they actually display.

We agree with this comment and have added the explanation of the dimension letters in the caption of the figures, with R for realizations (ensemble members), T for time, M for model, and S for scenarios.

5. Showing the interactions is of interest but I completely miss a physical picture of how I should interpret these interactions and what physical process they represent. What does it mean? Please explain in the text.

As mentioned in Section 2.2.1, "*interaction occurs when the separate effects of the factors do not combine additively*." These interactions therefore capture the fraction of variance that is inherently located in two or more dimensions. We acknowledge the reviewer's comment regarding the lack of physical interpretation of these interaction terms in the manuscript. In response, further analysis based on a minimalist synthetic model of AMOC trajectories (detailed in Appendix A) has allowed us to better clarify their physical significance. We show that interactions involving the ensemble dimensions (e.g., SR, RT, SRT …) represent internal variability. Meanwhile, the interaction between scenario and time (ST) reflects the portion of scenario-driven variability that is smoothed by the rolling time window. To clarify and introduce the interpretation of ANOVA components, we have now added a dedicated subsection, "***Interpretation and combination of ANOVA components***" (Section 2.2.2), to the methodology to support clearer understanding.

6. I would rephrase the conclusion between anthropogenic forcing intensity (which could be positive as well in terms of AMOC: aerosols!) and ensemble variance decrease or decrease of natural variability by saying there is a strong link between forced AMOC weakening and decrease in ….(l 318)

We agree with this suggestion and have updated the text accordingly:

"***These results therefore suggest a strong link between forced AMOC weakening and decrease in the AMOC ensemble variance.***" (l. 407-408)

7. Discussion on line 319-320 and further down in 3.2.2. I see what you say in your figures, but don't have a clear picture of what SRT interaction actually means in terms of physical processes, so this part is not meaningful to me (see also point 5).

In agreement with point 5, we agree with the reviewer that we lack a physical interpretation of the interactions. We have updated the manuscript with a dedicated subsection, "***Interpretation and combination of ANOVA components***" (Section 2.2.2), to the methodology to support clearer understanding.

8. My physical interpretation of 2.3 is that up to 2050 scenarios with stronger greenhouse forcing have smaller weakening in aerosol emissions and vice versa, making the impact on AMOC almost scenario independent, until the aerosol effect is gone, and we see the effect of greenhouse gas forcing alone.

The reviewer seems to suggest that the delayed emergence of the main scenario effect on variability is associated with the compensating effect of greenhouse gas (GHG) emissions and aerosol forcing. The SSP1-2.6 scenario shows a smaller increase in GHGs and a larger decrease in aerosols, and vice versa for the SSP5-8.5 scenario. However, this is not consistent with the evolution of the net radiative forcing imbalance (which account for both GHG and aerosol), which shows a divergence of the scenarios rapidly after their start in 2015 (O'Neill et al., 2016).

9. Discussion around line 360. I disagree with the interpretation, see point 3.

We agree with the reviewer and have updated the manuscript to explain that this difference between small and large ensembles regarding model main effect (M) is not associated with the size of the ensembles but with the individual characteristics of the models selected in each category.

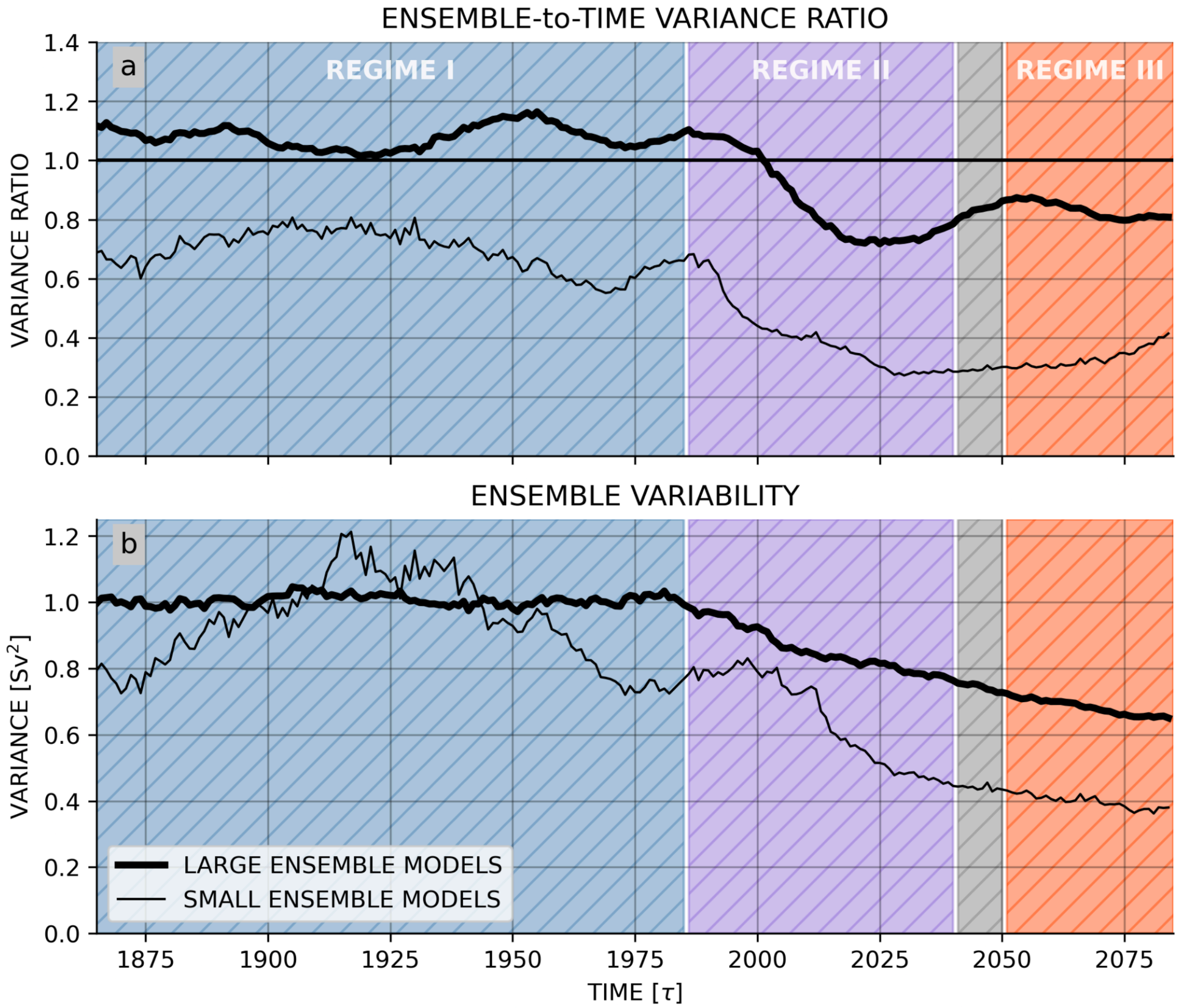
"***This is not due to any inherent characteristic of small ensembles, but rather associated with the particular combination of models in the "small ensemble" and***

"large ensemble" category, which happen to exhibit different AMOC sensitivities to forcing” (l. 448-450)

10. Summary your point 2. I think what you see in phase 2 is also forced by SSP scenarios, but as said before the net forcing on AMOC is reasonably equal for the scenarios. SSP126 forces stronger AMOC weakening than SSP245 and SSP585 in the first half of this century because the aerosols are faster removed from the atmosphere in this scenario.

In accordance with our response to point 8, this interpretation does not seem compatible with the trajectories of net radiative forcing imbalance in CMIP6, which separates between scenarios rapidly after 2015.

FIGURE 4 - UPDATED



References:

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geosci. Model Dev., 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.